SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
April 1998

3. REPORT TYPE AND DATES COVERED
Technical Report; April 1998

4. TITLE AND SUBTITLE
Asymptotic Distribution of the Random Regret Risk for Selecting Exponential Populations

5. FUNDING NUMBERS
DAAH04-95-1-0165

6. AUTHOR(S)
Shanti S. Gupta and Friedrich Liese

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Purdue University
Department of Statistics
West Lafayette IN 47907-1399

8. PERFORMING ORGANIZATION REPORT NUMBER
Technical Report #98-08C

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 32922.16-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
In this paper empirical Bayes methods are applied to construct selection rules for the selection of all good exponential distributions. We modify the selection rule introduced and studied by Gupta and Liang (1996) who proved that the regret risk $\mathbb{E}R_n$ converges to zero with rate $0(n^{-\lambda/2}), 0 < \lambda \leq 2$. The aim of this paper is to prove a limit theorem for the random regret risk R_n. It is shown that nR_n tends in distribution to a linear combination of independent χ^2- distributed random variables. This result especially implies that under weak conditions the random regret risk is of order $O_P(\frac{1}{n})$.

19981229 128

14. SUBJECT TERMS
Good exponential population, random regret risk, limit theorem, M-estimator

15. NUMBER IF PAGES
22

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Enclosure 1

ASYMPTOTIC DISTRIBUTION OF THE RANDOM REGRET RISK FOR SELECTING EXPONENTIAL POPULATIONS*

by

Shanti S. Gupta
Purdue University

and

Friedrich Liese[1]
University of Rostock
and Purdue University

Abstract

In this paper empirical Bayes methods are applied to construct selection rules for the selection of all good exponential distributions. We modify the selection rule introduced and studied by Gupta and Liang (1996) who proved that the regret risk $\mathbb{E}R_n$ converges to zero with rate $0(n^{-\lambda/2}), 0 < \lambda \leq 2$. The aim of this paper is to prove a limit theorem for the random regret risk R_n. It is shown that nR_n tends in distribution to a linear combination of independent χ^2- distributed random variables. This result especially implies that under weak conditions the random regret risk is of order $O_P(\frac{1}{n})$.

Short Title: Asymptotic distribution of Random Regret Risk

AMS 1991 Subject Classification: 62F07, 62C12

Key Words and Phrases: Good exponential populations, random regret risk, limit theorem, $M-$estimator.

---

*This research was supported in part by US Army Research Office, Grant DAAH04-95-1-0165 at Purdue University.

[1]This research was done while this author was a Visiting Professor at the Department of Statistics, Purdue University.

1. Introduction

The family of exponential distributions has fundamental meaning in reliability theory, survival analysis and general in the area of life time distributions. For an overview and more details we refer to Johnson, Kotz and Balakrishnan (1994) and Balakrishnan and Basu (1995). We consider k independent exponential populations $\pi_1, \ldots, \pi_k$ with expectations $\theta_1, \ldots, \theta_k$ which are unknown. Let there be a control value θ_0. Each population π_i is called good if $\theta_i \geq \theta_0$ and bad otherwise. We study the problem of finding all good populations. This is a typical subset selection problem, see Gupta and Panchapakesan (1985). We assume that the θ_i are random and distributed according to the unknown distribution G_i. Then for a given loss function the best selection rule, being the Bayes selection rule, depends on the unknown G_i. We suppose that historical data are available and can be included in the decision rule. This is the empirical Bayes approach due to Robbins (1956). Empirical Bayes methods have been applied in different areas of statistics. Deely (1965) constructed empirical Bayes subset selection procedures. In a series of papers Gupta and Liang (1988, 1994) and Gupta, Liang and Rau (1994a, 1994b) have studied different selection procedures using empirical Bayes approach.

Assume $\underline{Y}$ are the actual data based on which we wish to make a decision. Then the optimal decision d_G depends on the unknown joint distribution $G = \prod_{i=1}^{k} G_i$ of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$. The central idea of the empirical Bayes approach is the construction of a good decision rule d_n^* on the basis of historical data $\underline{Y}_n$. The quality of d_n^* is then characterized by the non-negative random regret risk $R_n^* = R(d_n^*, G) - R(d_G, G)$. The aim of the above mentioned papers dealing with empirical Bayes methods was to construct suitable decision rules d_n^* and to evaluate the non-random regret risk $\mathbb{E}R_n^*$. The main goal of these papers was to prove the convergence of $\mathbb{E}R_n^*$ to zero with a certain rate. Gupta and Liang (1996) constructed for the problem of selecting good exponential populations a selection rule d_n^* and proved $\mathbb{E}R_n^* = O(n^{-\lambda/2})$ with same $0 < \lambda \leq 2$.

The natural question is whether there exist other selection procedures which are possibly better in asymptotic sense. But if $\mathbb{E}R_n^* = O(n^{-1})$ then a comparison with another sequence d_n of selection rules would lead to the constants

$$\lim_{n\to\infty} n\mathbb{E}[R(d_n^*, G) - R(d_G, G)]$$

$$\lim_{n\to\infty} n\mathbb{E}[R(\hat{d}_n, G) - R(d_G, G)]$$

which have to be calculated and compared. At least two reasons are against this argument. Even if the stochastic order of R_n^* is $O_P(\frac{1}{n})$, the regret risk $\mathbb{E}R_n^*$ may not have the order $O(\frac{1}{n})$ as the values of R_n^* can be large on some events with probability tending to zero. As these events do not occur in most cases, the random regret risk better reflects the situation in which the empirical Bayes methods are applied. These methods behave better than indicated by the order of the regret risk $\mathbb{E}R_n^*$. The situation is comparable with the asymptotic theory of parameter estimation where different types of estimators are compared by the limit distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ and in general not by a direct evaluation of the variance of $\widehat{\theta}_n$. A second more technical argument is that $\lim_{n\to\infty} n\mathbb{E}[R(d_n^*, G) - R(d_G, G)]$ can be calculated only in very special situations. So, in this paper we study distributions instead of expectations. A new selection rule $\hat{d}_n$ for the problem of selecting a good exponential distribution is introduced by a modification of the Gupta, Liang rule (1996). We show

$$n[R(\hat{d}_n, G) - R(d_G, G)]$$

converges in distributions to a linear combination of independent χ^2−distributed random variables each with one degree of freedom. The coefficients in the linear combination are explicitly calculated. The main idea for the new selection rule is the fact that the construction of the optimal selection rule d_G needs only the value of a unique zero η_{i0} of some function H_i which depends on G_i. The main part of this paper is the construction of a suitable estimator $\widehat{\eta}_{in}$ for η_{i0} and the proof of a limit theorem for $\sqrt{n}(\widehat{\eta}_{in} - \eta_{i0})$.

2. Formulation of the Selection Problem

Consider k independent exponential populations $\pi_1, \ldots, \pi_k$ which we assumed to have the density functions $h(x_i|\theta_i) = I(x_i \geq 0)\frac{1}{\theta_i}\exp\{-\frac{x_i}{\theta_i}\}$, $i = 1, \ldots, k$, where $\underline{\theta} = (\theta_1, \ldots, \theta_k)\epsilon\Omega = (0, \infty)^k$ and $I(A)$ is the indicator function of the set A. Given a standard value $\theta_0 > 0$ we call a population π_j good if $\theta_j \geq \theta_0$. Our aim is to select all good populations. Therefore the decision space is $\mathbb{D} = \{0, 1\}^k = \{(a_1, \ldots, a_k) : a_i\epsilon\{0, 1\}\}$ and π_i is selected if and only if $a_i = 1$. Similar as in Gupta, Liang (1996), Gupta, Liang (1994) we use the loss function

$$L(\underline{\theta}, \underline{a}) = \sum_{i=1}^{k} \ell(\theta_i, a_i)$$

where

$$\ell(\theta_i, a_i) = a_i \theta_i (\theta_0 - \theta_i) I(0 < \theta_i < \theta_0) + (1 - a_i)\theta_i(\theta_i - \theta_0) I(\theta_0 \leq \theta_i)$$

By a selection rule $d = (a_1, \ldots, a_k)$ we shall mean a measurable mapping of the sample space $\mathcal{Y} = (0, \infty)^k$ into the decision space $\mathbb{D}$. If we have a measurement Y_i from each population π_i then the risk of the selection rule d is given by

$$R(\underline{\theta}, d) = \mathbb{E}L(\underline{\theta}, d(\underline{Y}))$$

where $\underline{Y} = (Y_1, \ldots, Y_k)$. In terms of densities the risk $R(\underline{\theta}, d)$ is also given by

$$R(\underline{\theta}, d) = \int L(\underline{\theta}, d(\underline{y}))h(\underline{y}, \underline{\theta})d\underline{y}$$

where $h(\underline{y}|\underline{\theta}) = \prod_{i=1}^{k} h(y_i|\theta_i)$ and $d\underline{y} = dy_1, \ldots dy_k$.

Using the loss function (2.1) we get

$$R(\underline{\theta}), d) = \sum_{i=1}^{k} \mathbb{E}\ell(\theta_i, q_i(Yi)) \tag{2.1}$$

where

$$q_i(y_i) = \mathbb{E}a_i(Y_1, \ldots, Y_{i-1}, y_i, Y_{i+1}, \ldots, Y_k) \tag{2.2}$$

The formula (2.1) shows that due to the special structure of the loss function we may restrict ourselves to randomized decisions q_i which depend on the data of π_i only. Further

$$\mathbb{E}\ell(\theta_i, q_i(Y_i)) = \int_0^{\infty} q_i(y_i)(\theta_0 - \theta_i)\ e^{-\frac{y_i}{\theta_i}} dy_i + C_i\ (\theta_i) \tag{2.3}$$

where $C_i(\theta_i) = \theta_i(\theta_i - \theta_0)I(\theta_i \geq \theta_0)$. As we will apply the Bayes and the empirical Bayes approach to the selection problem we assume that the θ_i are realizations of independent random variables Θ_i. The Θ_i are assumed to take values in $(0, \infty)$ and have distribution

G_i. The distribution G of the random vector $\underline{\Theta} = (\Theta_1, \ldots, \Theta_k)$ is then the product of the G_i. Furthermore $h(\underline{y}|\underline{\theta})$ is the conditional density of $\underline{Y}$ given $\underline{\Theta} = \underline{\theta}$.

We suppose

$$\int_0^\infty \theta_i^2 \, dG_i(\theta_i) < \infty, \; i = 1, \ldots, k \tag{2.4}$$

Then the Bayes risk is finite and given by

$$R(G, d) = \mathbb{E}L(\underline{\Theta}), d(\underline{Y}))$$

Using (2.3) we get

$$R(G,d) = \sum_{i=1}^{k} \int_0^\infty \int_0^\infty q_i(y_i)(\theta_0 - \theta_i)e^{-\frac{y_i}{\theta_i}} \, dG_i(\theta_i)dy_i + \gamma_i$$

where

$$\gamma_i = \int_0^\infty \theta_i(\theta_i - \theta_0)dG_i(\theta_i).$$

As in Gupta, Liang (1996) one obtains by integration by parts

$$(\theta_0 - \theta_i)e^{-\frac{y_i}{\theta_i}} = \int_{y_i}^\infty (\theta_0 + y_i - t_i)\frac{1}{\theta_i}e^{-\frac{t_i}{\theta_i}} dt_i$$

and

$$\int_0^\infty (\theta_0 - \theta_i)e^{-\frac{y_i}{\theta_i}} dG_i(\theta_i) = \theta_0\psi_{1i}(y_i) - \psi_{2i}(y_i)$$

where

$$\begin{aligned}\psi_{i1}(y_i) &= \int_0^\infty e^{-\frac{y_i}{\theta_i}} dG_i(\theta_i), \\ &= \int_0^\infty \int_{y_i}^\infty \frac{1}{\theta_i}e^{-\frac{t_i}{\theta_i}} dt_i \; dG_i(\theta_i) = \mathbb{E}(Y_i > y_i)\end{aligned} \tag{2.5}$$

$$\begin{aligned}\psi_{i2}(y_i) &= \int_0^\infty \theta_i e^{-\frac{y_i}{\theta_i}} dG_i(\theta_i), \\ &= \int_0^\infty \int_{y_i}^\infty (t_i - y_i)\frac{1}{\theta_i}e^{-\frac{t_i}{\theta_i}} dt_i dG_i(\theta_i) \\ &= \mathbb{E}(Y_i - y_i)I(Y_i > y_i)\end{aligned} \tag{2.6}$$

Using these relations we obtain

$$R(G,d) = \sum_{i=1}^{k} \int_0^\infty (\theta_0 \psi_{i1}(y_i) - \psi_{i2}(y_i)) q_i(y_i) dy_i + \gamma_i$$

where γ_i is independent of the selection procedure. This shows that $\inf_d R(G,d)$ is attained by the selection rule $d^0 = (d_1^0, \ldots, d_k^0)$ where

$$d_i^0(y_i) = \begin{cases} 1 & \text{if} \quad \theta_0 \psi_{i1}(y_i) \le \psi_{i2}(y_i) \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

If G_i is nondegenerate then $\frac{\psi_{i2}}{\psi_{i1}}$ is strictly increasing. This means that the zero η_{i0} of $\theta_0 \psi_{i1} - \psi_{i2}$, if there is any, is uniquely determined. To apply the selection rule d^0 we have to know the η_{i0}. But the ψ_{1i}, ψ_{2i} as well as η_{i0} include the unknown prior distribution. Otherwise the unknown ψ_{i1}, ψ_{i2} are the expectation of some functions of the observable data Y_i. This is the key to apply empirical Bayes methods. Assume we have data from the past which can be taken into consideration. More precisely let $Y_{i1}, \ldots, Y_{in}, i = 1, \ldots, k$ be n i.i.d. random variables with density

$$f_i(y_i) = \int_0^\infty \frac{1}{\theta_i} e^{-\frac{y_i}{\theta_i}} dG_i(\theta_i)$$

The relations (2.5) and (2.6) show that

$$\widehat{H}_{in}(y) = \frac{1}{n} \sum_{\ell=1}^{n} (\theta_0 + y - Y_{i\ell}) I(Y_{i\ell} > y) \tag{2.8}$$

is an unbiased and consistent estimator for the unknown function $H_i(y) = \theta_0 \psi_{i1}(y) - \psi_{i2}(y)$. Using this fact Gupta and Liang (1996) introduced an empirical Bayes selection procedure d_n^* by setting

$$d_{in}^*(y_i) = \begin{cases} 1 & \text{if} \quad \widehat{H}_{in}(y_i) \le 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

3. Results

The Bayes risk of the empirical Bayes selection procedure d_n^* is

$$R(G, d_n^*) = \mathbb{E} \sum_{i=1}^{k} \int_0^\infty d_{in}^*(y_i) [\theta_0 \psi_{i1}(y_i) - \psi_{i2}(y_i)] dy_i + \gamma_i$$

The regret risk of the selection procedure is then given by

$$R(G, d_n^*) - R(G, d_n^0) = \mathbb{E} \sum_{i=1}^{k} \int_0^\infty [I(\widehat{H}_{in}(y_i) \le 0) - I(H_i(y_i) \le 0)] H_i(y_i) dy_i$$

Gupta and Liang (1996) studied the rate of convergence to zero of the above nonnegative difference and proved

$$R(G, d_n^*) - R(G, d_n^0) = O(n^{-\frac{\lambda}{2}}) \quad \text{for some} \quad 0 < \lambda \le 2.$$

In this paper we will study the random part of the regret risk. We will not directly deal with the decision rule d_n^*. According to (2.7) the essential part of the optimal selection rule d_i^0 is the zero η_{i0} of the function $H_i(y) = \theta_0 \psi_{i1}(y) - \psi_{i2}(y)$. The function $\frac{\psi_{i2}}{\psi_{i1}}$ is strictly increasing if G_i is non-degenerated. To guarantee that H_i has a zero we use the following Assumption A which was already introduced in Gupta, Liang (1996).

Assumption A: *It holds* $\lim_{y \downarrow 0} \frac{\psi_{i2}(y)}{\psi_{i1}(y)} < \theta_0 < \lim_{y \uparrow \infty} \frac{\psi_{i2}(y)}{\psi_{i1}(y)}$.

If Assumption A is fulfilled and η_{i0} is the uniquely determined zero of H_i then

$$H_i(y) > 0 \quad \text{for} \quad y < \eta_{i0} \quad \text{and} \quad H_i(y) < 0 \quad \text{for} \quad y > \eta_{i0}.$$

We will construct a consistent estimator $\widehat{\eta}_{in}$ for η_{i0} and set

(3.1)
$$\widehat{d}_{in}(y_i) = \begin{cases} 1 & y_i \ge \widehat{\eta}_{in} \\ 0 & \text{otherwise} \end{cases}$$

Put

$$M_i(y) = \int_0^y H_i(t) dt$$

Then the regret risk of $\widehat{d}_n = (\widehat{d}_{n1}, \ldots, \widehat{d}_{nk})$ is

(3.2)
$$\mathbb{E} \sum_{i=1}^{k} [M_i(\eta_{i0}) - M_i(\widehat{\eta}_{in})] = \mathbb{E} R_n$$

where $R_n = \sum_{i=1}^{k} [M_i(\eta_{in} - M_i(\widehat{\eta}_{i0})]$ is the *random regret risk.* To prove a limit theorem for the random regret risk R_n we need at first a $\sqrt{n}-$ consistent estimator $\widehat{\eta}_{in}$ for η_{i0}. As the functions $\widehat{H}_{in}(y_i)$ are discontinuous the estimator for the zero of H_i can not be directly introduced as a zero of $\widehat{H}_{in}(y_i)$ as the existence of such zeroes is unclear. To get a continuous function we integrate $\widehat{H}_{in}$ and set

$$\widehat{M}_{in}(y_i) = \int_0^{y_i} \widehat{H}_{in}(s)ds \tag{3.3}$$

As $\widehat{H}_{in}(s)$ vanishes for all sufficiently large s we see that $M_{in}(y_i)$ is constant for all sufficiently large y_i. Taking into account the continuity of $\widehat{M}_{in}$ we get the existence of at least one maximum point, say $\widehat{\eta}_{in}$. From the law of large numbers we get that $\widehat{H}_{in}(s) \xrightarrow[n\to\infty]{P} H_i(s)$ and $\widehat{M}_{in}(y) \xrightarrow[n\to\infty]{P} M_i(y)$, where $\xrightarrow{P}$ denotes stochastic convergence. From the inclusion

$$\{| \widehat{\eta}_{in} - \eta_{i0} | \geq \epsilon\} \subseteq \{ \sup_{|y-\eta_{i0}|\geq\epsilon} \widehat{M}_{in}(y) \geq \widehat{M}_{in}(\eta_{i0})\} \tag{3.4}$$

one can see that in case of a unique maximum of M_i at η_{i0} and a uniform law of large numbers, i.e.

$$\sup_y | \widehat{M}_{in}(y) - M_i(y) | \xrightarrow[n\to\infty]{P} > 0$$

the $\widehat{\eta}_{in}$ will converge to η_{i0}. Starting with Wald (1949) this concept to prove consistency of estimators defined by minimization or maximation procedures have been used by many authors (see Pfanzagl (1969), Liese, Vajda (1994), van der Vaart, Wellner (1996) and the references there).

Theorem 1: *Suppose G_i is nondegenerate and (2.4) holds. Suppose Assumption A is fulfilled and $\widehat{\eta}_{in} \epsilon$ argmax $\widehat{M}_{in}$, where $\widehat{M}_{in}$ and $\widehat{H}_{in}$ are defined in (3.3) and (2.8), respectively. Then*

$$\widehat{\eta}_{in} \xrightarrow[n\to\infty]{P} \eta_{i0}$$

where η_{i0} is the uniquely determined zero of $H_i(y) = \theta_0\psi_{i1}(y) - \psi_{i2}(y)$.

The traditional way to prove asymptotic normality of estimators defined by a minimization procedure is to get an equation for this estimator by taking the derivative. The next step is a linearization of the obtained equation by Taylor expansion. This approach fails in our situation as the function M_{in} is not differentiable. But the one sided derivatives exist. The derivative from the right $D^+\widehat{M}_{in}(y_i)$ exists for every $y_i \geq 0$ and the derivative from the left $D^-\widehat{M}_{in}(y_i)$ exists for every $y_i > 0$. If $\widehat{\eta}_{in}$ is a maximum point of $\widehat{M}_{in}$ we get

$$D^+\widehat{M}_{in}(\widehat{\eta}_{in}) = \frac{1}{n}\sum_{\ell=1}^{n}(\theta_0 + \widehat{\eta}_{i1} - Y_{i\ell})I(Y_{i\ell} > \widehat{\eta}_{in})$$

$$= \widehat{H}_{in}(\widehat{\eta}_{in}) \leq 0$$

and for $\widehat{\eta}_{in} > 0$

$$D^-\widehat{M}_{in}(\widehat{\eta}_{in}) = \frac{1}{n}\sum_{\ell=1}^{n}(\theta_0 + (\widehat{\eta}_{in} - Y_{i\ell})I(Y_{i\ell} \geq \widehat{\eta}_{in})$$

$$= \widehat{H}_{in}(\widehat{\eta}_{in} - 0) \geq 0$$

Note that

$$D^+\widehat{M}_{in}(\widehat{\eta}_{in}) - D^-\widehat{M}_{in}(\widehat{\eta}_{in}) = \frac{\theta_0}{n}\sum_{\ell=1}^{n} I(Y_{i\ell} = \widehat{\eta}_{in}) \tag{3.5}$$

If the $Y_{i\ell}$ have continuous distributions then with probability one at most one term in the sum of the right hand side of (3.5) is nonzero. Denote by $Y_{i(1)} < Y_{i(2)} < \ldots < Y_{i(n)}$ the order statistics. If $\widehat{\eta}_{in} = Y_{i(r-1)}$ then for $Y_{i(r-1)} < y < Y_{i(r)}$

$$D^+\widehat{M}_{in}(y) = \frac{1}{n}\sum_{\ell=1}^{n}(\theta_0 + y - Y_{i\ell})I(Y_{i\ell} > y)$$

$$= \frac{1}{n}\sum_{\ell=r}^{n}(\theta_0 + y - Y_{i(\ell)})$$

$$> \frac{1}{n}\sum_{\ell=r}^{n}(\theta_0 + Y_{i(r-1)} - Y_{i(\ell)})$$

$$\geq D^+\widehat{M}_{in}(Y_{i(r-1)}) \geq 0$$

which contradicts the maximum property of $\widehat{\eta}_{in}$. Hence on the event $\{\widehat{\eta}_{in} > 0\}$ the right hand term in (3.5) is *a.s.* zero. Consequently

$$\widehat{H}_{in}(\widehat{\eta}_{in})I(\widehat{\eta}_{in} > 0) = 0 \quad a.s. \tag{3.6}$$

The function $\widehat{H}_{in}(y)$ is discontinuous. This excludes a linearization of (3.6) by Taylor expansion. But for large n we have $\widehat{H}_{in}(y) \approx \mathbb{E}\widehat{H}_{i1} = \theta_0\psi_{i1}(y) - \psi_{i2}(y)$. As $\theta_0\psi_{i1}(y) - \psi_{i2}(y)$ is smooth we will be able to derive an asymptotic linearization. The idea of an asymptotic linearization goes back to Pollard (1990) and was systematically used in Jurečkova and Sen (1996) to deal with regression models.

Denote by the variance of the $r.v. X$ by $V(X)$. Using integration by parts one can show that (2.4) implies $EY_i^2 < \infty$. We set

$$\sigma_{io}^2 = V((\theta_0 + \eta_{io} - Y_{i\ell})I(Y_{i\ell} \geq \eta_{i0})) \tag{3.7}$$

Denote by $N(\mu, \sigma^2)$ the distribution with expectation μ and variance σ^2.

Theorem 2: *Assume the conditions in Theorem 1 are fulfilled and it holds*

$$b_i = -[\theta_0 \psi'_{i1}(\eta_{i0}) - \psi'_{i2}(\eta_{i0})] > 0 \tag{3.8}$$

then

$$\sqrt{n}(\widehat{\eta}_{in} - \eta_{i0}) \xrightarrow[n\to\infty]{d} N(0, \sigma_{i0}^2/b_i^2)$$

where $\xrightarrow{d}$ *denotes convergence in distributions.*

Now we apply Theorem 2 to evaluate the random part R_n in the regret risk $\mathbb{E}R_n$. It holds

$$R_n = \sum_{i=1}^{k}[M_i(\eta_{i0}) - M_i(\widehat{\eta}_{in})]$$

Note that $M_i'(\eta_{i0}) = 0$ and by assumption (3.8) $-M_i''(\eta_{i0}) = -H_i'(\eta_{i0}) = b_i > 0$. Hence by Theorem 2

$$nR_n = \sum_{i=1}^{k} \frac{n}{2}(\widehat{\eta}_{in} - \eta_{i0})^2[b_i + o_P(1)] \tag{3.9}$$

Let $\mathcal{L}(X)$ denote the distribution of the random variables X. We denote by $\chi_1^2, \dots, \chi_k^2$ i.i.d. random variables whose common distribution is a χ^2- distribution with one degree of freedom. Then we get from (3.9) Theorem 2 and the Slutzky theorem the following result.

Theorem 3: *If the assumptions in Theorem 2 are fulfilled ,* $\hat{d}_{in}$ *is defined in (3.1) and*

$$R_n = \sum_{i=1}^{k}\int_0^\infty \widehat{d}_{in}(y_i)H_i(y_i)dy_i - \sum_{i=1}^{k}\int_0^\infty d_i^0(y_i)H_i(y_i)dy_i$$

is the random part in the regret risk (3.2), then

$$nR_n \xrightarrow[n\to\infty]{d} \mathcal{L}\Big(\sum_{i=1}^{k}\Big(\frac{\sigma_{i0}^2}{2b_i}\Big)\ \chi_i^2\Big)$$

4. Proofs:

As G_i is nondegenerate, the function $\frac{\psi_{i2}}{\psi_{i1}}$ continuous and is strictly increasing. Hence by Assumption A the function $H_i(y)$ has a uniquely determined zero η_{i0} and the following holds

$$H_i(y) > 0 \text{ for } y < \eta_{i0} \text{ and } H_i(y) < 0 \text{ for } y > \eta_{i0} \quad .$$

Consequently

$$M_i(y) = \int_0^y H_i(s)ds$$

has a uniquely determined maximum of $y = \eta_{i0}$. If we know that $\widehat{M}_{in}$ converges uniformly to M_i then we will expect that the maximum points of $\widehat{M}_{in}$ being the $\widehat{\eta}_{in}$ will tend to the minimum point η_{i0} of M_i. More precisely

$$\begin{aligned} \{|\ \widehat{\eta}_{in} - \eta_{i0}|\ > \epsilon\} &\subseteq \{ \sup_{y:|y-\eta_{i0}|>\epsilon} \widehat{M}_{in}(y) \geq \widehat{M}_{in}(\eta_{i0})\} \\ &\subseteq \{ \sup_{y:|y-\eta_{i0}|>\epsilon} (\widehat{M}_{in}(y) - \widehat{M}_{in}(\eta_{i0})) \geq 0\} \end{aligned} \tag{4.1}$$

Also

$$\sup_{y:|y-\eta_{i0}|>\epsilon} (\widehat{M}_{in}(y) - \widehat{M}_{in}(\eta_{i0})) \leq$$

$$\sup_{y:|y-\eta_{i0}|>\epsilon} (M_i(y) - M_i(\eta_{i0})) + 2 \sup_{0\leq y<\infty} |\widehat{M}_{in}(y) - M_i(y)|$$

We set

$$\delta_\epsilon = \sup_{y:|y-\eta_{i0}|>\epsilon} (M_i(\eta_{i0}) - M_i(y)) \tag{4.2}$$

and note that $\delta_\epsilon > 0$ as M_i is strictly increasing for $y < \eta_{i0}$ and strictly decreasing for $y > \eta_{i0}$. The inclusion (4.1) implies

$$\{|\ \widehat{\eta}_{in} - \eta_{i0}| > \epsilon\} \subseteq \{ \sup_{0\leq y<\infty} |\widehat{M}_{in}(y) - M_i(y)| > \frac{1}{2}\delta_\epsilon\} \tag{4.3}$$

which shows that a uniform law of large numbers implies the consistency of $\widehat{\eta}_{in}$.

Lemma 1. *Assume G_i is nondegenerate and it holds (2.4) and Assumption A. If $\widehat{M}_{in}$ is defined by (3.3) and $\widehat{\eta}_{in} \epsilon$ argmax $\widehat{M}_{in}$ then*

$$\widehat{\eta}_{in} \xrightarrow[n\to\infty]{d} \eta_{i0}$$

Proof: Note that

$$
\begin{aligned}
\widehat{M}_{in}(y) &= \int_0^y \widehat{M}_{in}(s)ds \\
&= \int_0^y \frac{1}{n}\sum_{\ell=1}^{n}(\theta_0 + s - Y_{i\ell})I(Y_{i\ell} > s)ds \qquad (4.4)\\
&= \frac{1}{n}\sum_{\ell=1}^{n}(Y_{i\ell}\wedge y)(\theta_0 - Y_{i\ell} + \frac{1}{2}(Y_{i\ell}\wedge y))
\end{aligned}
$$

where $Y_{i\ell} \wedge y = \min(Y_{i\ell}, y)$. Consequently

$$
\begin{aligned}
\widehat{M}_{in}(\infty) &= \lim_{y\to\infty} \widehat{M}_{in}(y) \\
&= \frac{1}{n}\sum_{\ell=1}^{n} Y_{i\ell}(\theta_0 - \frac{1}{2}Y_{i\ell})
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E} \sup_{T\le y<\infty} |\widehat{M}_{in}(y) - \widehat{M}_{in}(\infty)| \le \\
&\frac{1}{n}\sum_{\ell=1}^{n}[\mathbb{E} \sup_{T\le y<\infty} \theta_0|(Y_{i\ell}\wedge y) - Y_{i\ell}| + \mathbb{E} \sup_{T\le y<\infty} Y_{i\ell}|(Y_{i\ell}\wedge y) - Y_{i\ell}| : \\
&\qquad + \mathbb{E} \sup_{T\le y<\infty} \frac{1}{2}\ |(Y_{i\ell}\wedge y)^2 - Y_{i\ell}^2)| \\
&\le \frac{1}{n}\sum_{\ell=1}^{n}[\mathbb{E}\ \theta_0(Y_{i\ell} - T)I(Y_{i\ell}\ge T) + \mathbb{E}\ Y_{i\ell}(Y_{i\ell} - T)I(Y_{i\ell}\ge T) \\
&\qquad + \frac{1}{2}\ \mathbb{E}(Y_{i\ell}^2 - T^2)I(Y_{i\ell}\ge T)] \\
&= \theta_0\mathbb{E}(Y_{i1} - T)I(Y_{i1}\ge T) + \mathbb{E}Y_{i1}(Y_{i1} - T)I(Y_{i1}\ge T) \qquad (4.5)\\
&\qquad + \frac{1}{2}\mathbb{E}(Y_{i1}^2 - T^2)I(Y_{i1}\ge T)
\end{aligned}
$$

The assumption (2.4) implies $\mathbb{E}Y_{in}^2 < \infty$ and the theorem of Lebesgue yields that

$$
\lim_{T\to\infty} \mathbb{E} \sup_{T\le y<\infty} \mid \widehat{M}_{in}(y) - \widehat{M}_{in}(\infty) \mid = \lim_{T\to\infty} \Delta_T = 0 \qquad (4.6)
$$

Now we study $\widehat{M}_{in}(y)$ for $0 \leq y \leq T$ where $T > 0$ is fixed. It holds

$$| \widehat{M}_{in}(y+h) - \widehat{M}_{in}(y) | = | \int_y^{y+h} \widehat{M}_{in}(s) ds |$$

$$\leq \int_y^{y+h} \frac{1}{n} \sum_{\ell=1}^{n} | \theta_0 + s - Y_{i\ell} | I(Y_{i\ell} \geq s) ds$$

$$\leq h \frac{1}{n} \sum_{\ell=1}^{n} (\theta_0 + (Y_{i\ell} + h) + Y_{i\ell})$$

Hence

$$\mathbb{E} \sup_{y,h: 0 \leq y \leq y+h \leq T} | \widehat{M}_{in}(y+h) - \widehat{M}_{in}(y) | \leq h(\theta_0 + h) + h \mathbb{E} Y_{in} \tag{4.7}$$

For any function $f : [0, T] \longrightarrow (-\infty, +\infty)$ we denote by $\omega_f(h)$ the modul of continuity defined by

$$\omega_{f,T}(h) = \sup_{y,h: 0 \leq y \leq y+h \leq T} | f(y+h) - f(y) |$$

Then for every $n = 1, 2, \ldots$

$$\sup_{0 \leq y \leq T} | \widehat{M}_{in}(y) - M_i(y) | \leq \max_{0 \leq k \leq m} | \widehat{M}_{in}(\frac{Tk}{m}) - M_i(\frac{Tk}{m}) |$$

$$+ \omega_{\widehat{M}in,T}(\frac{T}{m}) + \omega_{Mi,T}(\frac{T}{m})$$

Hence

$$\begin{aligned} \mathbb{E} \sup_{0 \leq y \leq T} | \widehat{M}_{in}(y) - M_i(y) | &\leq \sum_{k=0}^{m} \mathbb{E} | \widehat{M}_{in}(\frac{Tk}{w}) - M_i(\frac{Tk}{w}) | \\ &+ \mathbb{E} \omega_{\widehat{M}in,T}(\frac{T}{m}) + \mathbb{E} \omega_{Mi,T}(\frac{T}{m}) \end{aligned} \tag{4.8}$$

and by (4.7), (4.8)

$$\begin{aligned}
\mathbb{E}\sup_{0\le y\le\infty} \mid \widehat{M}_{in}(y) - M_i(y) \mid \le \mathbb{E}\sup_{0\le y\le T} \mid \widehat{M}_{in}(y) - M_i(y) \mid \\
+\mathbb{E}\sup_{T\le y<\infty} \mid \widehat{M}_{in}(y) - \widehat{M}_{in}(\infty) \mid \\
+\mathbb{E} \mid \widehat{M}_{in}(\infty) - M_{in}(\infty) \mid \\
+\sup_{T\le y<\infty} |M_i(y) - M_i(\infty)) \mid \\
\le \sum_{k=0}^{m} \mathbb{E} \mid \widehat{M}_{in}(\frac{Tk}{m}) - M_i(\frac{Tk}{m}) \mid \\
+\frac{T}{m}(\theta_0 + \frac{T}{m}) + \frac{T}{m}\mathbb{E}Y_{in} \\
\omega_{\mathrm{M}_i,\mathrm{T}}(\frac{T}{m}) + \Delta_T + \mathbb{E} \mid \widehat{M}_{in}(\infty) - M_i(\infty) \mid
\end{aligned}$$

Given $\alpha > 0$ we may choose T_0 such that by (4.6)

$$\Delta_{T_0} < \frac{\alpha}{3} \tag{4.9}$$

Then we fix m_0 such that

$$\frac{T_0}{m_0}(\theta_0 + \frac{T_0}{m_0}) + \frac{T_0}{m_0}\ \mathbb{E}\ Y_{in} + \omega_{\mathrm{M}_i,\mathrm{T}_0}(\frac{T_0}{m_0}) < \frac{\alpha}{3} \tag{4.10}$$

Here we have used that M_i is uniformly continuous on $[0, T_0]$. By the law of large numbers we find n_0 such that for every $n > n_0$

$$\sum_{k=0}^{m_0} \mathbb{E} \mid \widehat{M}_{in}(\frac{T_0 k}{m_0}) - M_i(\frac{T_0 k}{m_0}) \mid + \mathbb{E} \mid \widehat{M}_{in}(\infty) - M_i(\infty) \mid < \frac{\alpha}{3} \tag{4.11}$$

The combination of (4.9), (4.10), (4.11) yields

$$\mathbb{E}\sup_{0\le y<\infty} \mid \widehat{M}_{in}(y) - M_i(y) \mid < \alpha$$

for every $n > n_0$. This yields

$$P(\{\sup_{0\le y<\infty} \mid \widehat{M}_{in}(y) - M_i(y) \mid > \frac{1}{2}\delta_\epsilon\}) \underset{n\to\infty}{\longrightarrow} 0$$

which implies the consistency of $\widehat{\eta}_{in}$ in view of (4.3).

The classical method to deal with consistent estimators which fulfill an equation is the linearization of the equation by Taylor expansion. However this technique is not applicable to (3.6) as the corresponding functions are not differentiable. They are even not continuous. We will use another approach and derive a more implicit representation by a stochastic process. To prepare this representation of $\widehat{\eta}_{in}$ we need some technical lemmas. The first is the result of straight forward calculations.

Lemma 2: *Let $(U_i, V_i), i = 1, \dots, n$ be i.i.d. random vectors with $\mathbb{E}U_i^2 < \infty, \mathbb{E}V_i^2 < \infty$ and $\mathbb{E}U_i = \mathbb{E}V_i = 0$. Then*

$$\begin{aligned} \mathbb{E}(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} U_k)^2(\frac{1}{\sqrt{n}}\sum_{\ell=1}^{n} V_\ell)^2 &= \frac{1}{n^2}\sum_{k\neq\ell}(\mathbb{E}U_k^2)(\mathbb{E}V_\ell^2) \\ +2\sum_{k\neq e}(\mathbb{E}U_k V_k)(\mathbb{E}U_\ell V_\ell) &+ \frac{1}{n^2}\sum_{\ell=1}^{n}\mathbb{E}U_\ell^2 V_\ell^2 \\ &\leq \frac{3(n-1)}{n}(\mathbb{E}U_1^2)(\mathbb{E}V_1^2) + \frac{1}{n}\mathbb{E}U_1^2 V_1^2 \end{aligned} \tag{4.12}$$

Set

$$A_{ij}(s) = (\theta_0 + s - Y_{ij})I(Y_{ij} > s) \tag{4.13}$$

We fix $0 < a < b < \infty$ and consider the stochastic processes $A_{ij}(s)$ in $[a, b]$. It holds for $a < s < t < b$

$$A_{ij}(t) - A_{ij}(s) = (Y_{ij} - \theta_0 - s)I(s < Y_{ij} \leq t) + (t - s)I(Y_{ij} > t) \tag{4.14}$$

Hence by $(\alpha + \beta)^2 \leq 2(\alpha^2 + \beta^2)$

$$\begin{aligned} &\mathbb{E}[(A_{ij}(t) - \mathbb{E}A_{ij}(t)) - (A_{ij}(s) - \mathbb{E}A_{ij}(s))]^2 \leq \mathbb{E}(A_{ij}(t) - A_{ij}(s))^2 \\ &\leq 2[\theta_0 + 2b)^2(F_i(t) - F_i(s)) + (t - s)] \end{aligned} \tag{4.15}$$

where F_i is the *c.d.f.* of $Y_{ij}, j = 1, \dots, n$.

Furthermore for $a \leq s < t < u < b$

$$(A_{ij})(t) - A_{ij}(s))(A_{ij}(u) - A_{ij}(t))$$

$$= [(Y_{ij} - \theta_0 - s)I(s < Y_{ij} \leq t) + (t-s)I(Y_{ij} > t)].$$

$$[(Y_{ij} - \theta_0 - t)I(t < Y_{ij} \leq u) + (u-t)I(Y_{ij} > u)]$$

$$= (t-s)(Y_{ij} - \theta_0 - t)I(t < Y_{ij} \leq u)$$

$$+ (t-s)(u-t)I(Y_{ij} > u)$$

Using again $(\alpha + \beta)^2 \leq 2(\alpha^2 + \beta^2)$ we obtain for any $r.v$ X, Y with finite fourth moment

$$\mathbb{E}(X - \mathbb{E}X)^2(Y - \mathbb{E}Y)^2 \leq 4\mathbb{E}(X^2 + (\mathbb{E}X)^2)(Y^2 + (\mathbb{E}Y)^2)$$

$$\leq 4(\mathbb{E}X^2Y^2 + (\mathbb{E}X)^2\mathbb{E}Y^2 + (\mathbb{E}Y)^2\mathbb{E}X^2 + (\mathbb{E}X)^2(EY)^2) \leq \ 16\mathbb{E}X^2Y^2$$

Putting

$$X = (t-s)(Y_{ij} - \theta_0 - t)I(t < Y_{ij} < u)$$

$$Y = (t-s)(u-t)I(Y_{ij} > u)$$

we arrive at

$$\mathbb{E}[(A_{ij}(t) - \mathbb{E}A_{ij}(t)) - (A_{ij}(s) - \mathbb{E}A_{ij}(s))]^2[(A_{ij}(u) - \mathbb{E}A_{ij}(u)) - (A_{ij}(t) - \mathbb{E}A_{ij}(t))]^2$$

$$\leq 16\ \mathbb{E}(t-s)^2[(Y_{ij} - \theta_0 - t)I(t < Y_{ij} < u) + (u-t)I(Y_{ij} \geq u)]^2$$

$$\leq 16\ (t-s)^2 2\mathbb{E}((Y_{ij} - \theta_0 - t)^2 I(t < Y_{ij} < u) + (u-t)^2)$$

$$\leq 16(t-s)^2 2((\theta_0 + 2b)^2 + b^2) \tag{4.16}$$

Recall that Y_{ij} has the density

$$f_i(y_i) = \int_0^\infty \frac{1}{\theta_i}\ e^{-\frac{y_i}{\theta_0}} dG_i(\theta_i)$$

We introduce the continuous nondecreasing function $K_i(t), a \leq t \leq b$, by

$$K_i(t) = \int_a^t D(f_i(s) + 1)ds \tag{4.17}$$

where $D = 32(\theta_0 + 2b)^2 + b^2 + 2$ Then by (4.15) and (4.16)

$$\mathbb{E}[(A_{ij}(t) - \mathbb{E}A_{ij}(t)) - (A_{ij}(s) - \mathbb{E}A_{ij}(s))]^2 \leq K_i(t) - K_i(s) \tag{4.18}$$

$$\mathbb{E}[(A_{ij}(t)-\mathbb{E}A_{ij}(t))-(A_{ij}(s)-\mathbb{E}A_{ij}(s))]^2[(A_{ij}(u)-\mathbb{E}A_{ij}(u))(A_{ij}(t)-\mathbb{E}A_{ij}(t))]^2$$

$$\leq (K_i(t)-K_i(s))^2. \tag{4.19}$$

Introduce $W_{in}(t), a \leq t \leq b$, by

$$W_{in}(t)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(A_{ij}(t)-\mathbb{E}A_{ij}(t))$$

Then by the definition of $\widehat{H}_{in}(t)$ and $H_i(t)$, we have

$$W_{in}(t)=\sqrt{n}\ (\widehat{H}_{in}(t)-H_i(t)) \tag{4.20}$$

For fixed $a \leq s < t < u \leq b$ we apply Lemma 2 to

$$U_j=(A_{ij}(t)-\mathbb{E}A_{ij}(t))-(A_{ij}(s)-\mathbb{E}A_{ij}(s))$$
$$V_j=(A_{ij}(u)-\mathbb{E}A_{ij}(u))-(A_{ij}(t)-\mathbb{E}A_{ij}(t))$$

to get

$$\mathbb{E}[(W_{in}(t)-W_{in}(s))(W_{in}(u)-W_{in}(t))]^2 \leq \frac{3(n-1)}{n}\mathbb{E}U_1^2V_1^2+\frac{1}{n}\mathbb{E}U_1^2V_1^2$$

The application of the inequalities (4.18), (4.19) to right hand terms yields

$$\mathbb{E}[(W_{in}(t)-W_{in}(s))(W_{in}(u)-W_{in}(t))]^2 \leq 3(K_i(u)-K_i(s))^2 \tag{4.21}$$

We have also to deal with the fourth moments of the process W_{in}. To this end we use the following well known formula for the fourth moment of a sum of i.i.d. random variables.

Lemma 3: *Let $Z_1,\ldots,Z_n$ be i.i.d. random variables with $\mathbb{E}Z_i=0, \sigma^2=\mathbb{E}Z_i^2$ and $\mu_4=\mathbb{E}Z_i^4<\infty$. Then*

$$\mathbb{E}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i)^4=3\frac{n-1}{u}\sigma^4+\frac{1}{n}\mu_4.$$

Now we set

$$Z_{ij}=(A_{ij}(t)-\mathbb{E}A_{ij}(t))-(A_{ij}(s)-\mathbb{E}A_{ij}(s))$$

and note that by (4.18)

$$\sigma^2=\mathbb{E}Z_{ij}^2 \leq K_i(t)-K_i(s)$$

The representation (4.14) shows that the *r.v.* Z_{ij} are bounded

$$| Z_{ij}| \leq 4(b - a + \theta_0)$$

Hence

$$\mu_4 \leq (4(b - a + \theta_0))^2 \sigma^2$$

and with $d = (4(b - a + \theta_0))^2$

$$\mu_4 \leq d(K_i(t) - K_i(s))$$

If we now apply Lemma 3 to W_{in} then

$$\mathbb{E}(W_{in}(t) - W_{in}(s))^4 \leq 3(K_i(t) - K_i(s))^2 + \frac{d}{n}(K_i(t) - K_i(s)) \tag{4.22}$$

Recall that the model of continuity $\omega_f(h)$ of a function of $[a, b]$ is defined by

$$\omega_f(h) = \sup_{a \leq t \leq t+h \leq b} | (f(t + h) - f(t) |$$

For the proof of the next lemma we refer to Shorack and Wellner (1986), p. 49. Suppose $Z(t), 0, \leq t \leq 1$ is a stochastic process for which every path is continuous from right and has limits from the left.

Lemma 4: *Assume there is a continuous nondecreasing function K on $[0, 1]$ such that*

$$\mathbb{E}(Z(t) - Z(s))^2(Z(u) - Z(t))^2 \leq (K(u) - K(t))^2 \tag{4.23}$$

for every $0 \leq s \leq t \leq u \leq 1$. Then there is a universal constant c such that

$$\begin{aligned} P(\omega_Z(\frac{1}{m}) \geq \epsilon) &\leq \frac{1}{\epsilon^4} \sum_{k=1}^{m} \mathbb{E}(Z(\frac{k}{m}) - Z(\frac{k-1}{m}))^4 \\ &+ \frac{c(K(1) - K(0))}{\epsilon^4} \omega_K(\frac{1}{m}) \end{aligned} \tag{4.24}$$

To apply Lemma 4 to W_{in} we set

$$Z(t) = W_{in}(a + t(b - a))$$

$$K(t) = \sqrt{3} K_i(a + t(b - a))$$

In view of (4.21) the condition (4.23) is fulfilled. Note that by the definition of Z and inequality (4.22)

$$\sum_{k=1}^{m} \mathbb{E}(Z(\frac{k}{m}) - Z(\frac{k-1}{m}))^4$$

$$\leq \sum_{k=1}^{m} 3(K_i(a + \frac{(b-a)}{m}k) - K_i(a + \frac{(b-a)(k-1)}{m}))^2 + \frac{d}{n}K_i(b)$$

$$\leq (3\omega_{K_i}(\frac{b-a}{m}) + \frac{d}{n})K_i(b)$$

Consequently by (4.24)

$$P(\omega_{W_{in}}(\frac{b-a}{m}) \geq \epsilon) \leq (\frac{3+c}{\epsilon^4}\,\omega_{K_i}(\frac{b-a}{m}) + \frac{d}{\epsilon^4 n})K_i(b) \tag{4.25}$$

Now we are ready to prove an asymptotic representation for the estimator $\widehat{\eta}_{in}$.

Lemma 5: *Under the assumptions of Theorem 2 it holds*

$$\sqrt{n}\; H_i(\widehat{\eta}_{in}) =^- W_{in}(\eta_{i0}) + \rho_n$$

where

$$\rho_n \xrightarrow[n\to\infty]{P} 0.$$

Proof:

Fix $\theta < a < \eta_{i0} < b < \infty$. Then by the consistency of $\widehat{\eta}_{in}$ we get $P(\widehat{\eta}_{in}\epsilon[a,b]) \xrightarrow[n\to\infty]{} 1$. If $\widehat{\eta}_{in}\epsilon[a,b]$ then by (3.6) and (4.20)

$$0 = \widehat{H}_{in}(\widehat{\eta}_{in})\sqrt{n}$$

$$= W_{in}(\widehat{\eta}_{in}) + \sqrt{n}\; H_i(\widehat{\eta}_{in})$$

$$= W_{in}(\eta_{i0}) + \sqrt{n}\; H_i(\widehat{\eta}_{in}) + (W_{in}(\widehat{\eta}_{in}) - W_{in}(\eta_{i0}))$$

Introduce

$$C_{mn} = \sup_{|t-\eta_{i0}|\leq \frac{b-a}{m}} |W_{in}(t) - W_{in}(\eta_{i0})|$$

$$D_{mn} = \{|\widehat{\eta}_{in} - \eta_{i0}| \leq \frac{b-a}{m}\}$$

Then

$$I(D_{mn})[W_{in}(\eta_{i0}) + \sqrt{n}\ H_i(\widehat{\eta}_{in}) + (W_{in}(\widehat{\eta}_{in}) - W_{in}(\eta_{i0}))] = 0$$

or

$$W_{in}(\eta_{i0}) + \sqrt{n}\ H_i(\widehat{\eta}_{in}) = \rho_{1n} + \rho_{2n}$$

where

$$\rho_{1n} = I(\overline{D}_{mn})[W_{in}(\eta_{i0}) + \sqrt{n}\ H_i(\widehat{\eta}_{in})]$$

$$\rho_{2n} = I(D_{mn})(W_{in}(\widehat{\eta}_{in}) - W_{in}(\eta_{i0})$$

To show $\rho_{1n} \underset{n\to\infty}{\xrightarrow{P}} 0$ it suffices to remark that by the consistency of $\widehat{\eta}_{in}$:

$$P(\overline{D}_{mn}) \underset{n\to\infty}{\longrightarrow} 0$$

But by (4.25)

$$P(|\rho_{2n}| > \epsilon) \leq P(\omega_{W_{in}}(\frac{b-a}{m}) \geq \epsilon)$$

$$\leq (\frac{3+c}{\epsilon^4}\omega_{K_i}(\frac{b-a}{m}) + \frac{d}{\epsilon^4 n})K_i(b)$$

As K_i is continuous we see $\omega_{K_i}\left(\frac{b-a}{m}\right) \underset{m\to\infty}{\longrightarrow} 0$. Taking at first $n \to \infty$ and then $m \to \infty$ we see that

$$P(|\ \rho_{2n}\ | > \epsilon) \underset{n\to\infty}{\longrightarrow} 0$$

Hence $\rho_{1n} + \rho_{2n} \underset{n\to\infty}{\xrightarrow{P}} 0$ which proves the statement.

Proof of Theorem 2: As $H_i(\eta_{i0}) = 0$ and $H_i(y) = \theta_0\ \psi_{in}(y) - \psi_{in}(y)$ is differentiable at $y = \eta_{i0}$ we obtain from the consistency of $\widehat{\eta}_{in}$ and Lemma 5

$$\sqrt{n}\ (\widehat{\eta}_{in} - \eta_{i0})(H_i'(\eta_{i0}) + S_n) = -W_{in}(\eta_{i0}) + \rho_n$$

where both S_n and ρ_n tend stochastically to zero. By the central limit theorem the distribution of $W_{in}(\eta_{i0})$ tends to $N(0, \sigma_0^2)$ as $n \to \infty$. The application of the Slutzky Lemma yields the statement.

References

Balakrishnan, N. and Basu, A. P. (eds.) (1995). The Exponential Distribution: Theory, Method and Applications. Gordon and Breach Publishers. Langliorne, Pennsylvania.

Deely, J. J. (1965). Multiple decision procedures from empirical Bayes approach, Ph.D. Thesis (Mimeo. Ser. No. 45). Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Panchapakesan, S. (1985). Subset selection procedures: review and assessment. *Amer. J. Management Math. Sci.*,5, 235–311.

Gupta, S. S. and Liang T. (1986). Empirical Bayes rules for selecting the best binomial population. *Statistical Decision Theory and Related Topics IV* (Eds. S. S. Gupta and J. O. Berger) Vol. 1, 213–224, Springer-Verlag.

Gupta, S. S. and Liang, T. (1994). On empirical Bayes selection rules for sampling inspection. *J. Statist. Plann. Inference*,38, 43–64.

Gupta, S. S., Liang, T. and Rau, R.-B. (1994a). Empirical Bayes two stage procedures for selecting the best Bernoulli population compared with a control. *Statistical Decision Theory and Related Topics V.* (Eds. S. S. Gupta and J. O. Berger), 277–292, Springer Verlag.

Gupta, S. S., Liang, T. and Rau, R.-B. (1994b). Empirical Bayes rules for selecting the best normal population compared with a control. *Statistics &Decision*, 12, 125–147.

Gupta, S. S., Liang, T. (1996). Selecting good exponential populations compared with a control: nonparametric empirical Bayes approach, Technical Report #96-18C, Center for Statistical Decision Sciences and Department of Statistics.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). Continuous Univariate Distributions, Vol. 1, Second Ed., John Wiley & Sons, New York.

Jurečkova, J. and Sen, P. K. (1996). Robust Statistical Procedures, Asymptotics and Interrelations, J. Wiley & Sons, New York.

Liese, F. and Vajda, I. (1994). Consistency of M-estimates in general regression models. *Journ. Multiv. Analysis*,50, 93–114.

Pfanzagl, J. (1969). On measurability and consistency of minimum contrast estimators, *Metrika*, 14, 249–272.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7, 186–199.

Robbins, H. (1956). An empirical Bayes approach to statistics. Proc. Third Berkeley Symp., *Math. Statist. Probab.* 1, 157–163, Univ. of California Press.

Skorack, G. R. and Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. J. Wiley & Sons, New York.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Amer. Math. Stat.* 20, 595–601.

Van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes. Springer Series in Statistics, Springer Verlag.